working paper department of economics

Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings

Alberto Abadie
Joshua Angrist
Guido Imbens

No. 99-16 October 1999

massachusetts institute of technology
50 memorial drive
cambridge, mass. 02139

WORKING PAPER
DEPARTMENT
OF ECONOMICS

Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings

Alberto Abadie
Joshua Angrist
Guido Imbens

No. 99-16 October 1999

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

50 MEMORIAL DRIVE
CAMBRIDGE, MASS. 02142

Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings*

Alberto Abadie — MIT
Joshua Angrist — MIT and NBER
Guido Imbens — UCLA and NBER

Revised: September 1999

Abstract

The effect of government programs on the distribution of participants' earnings is important for program evaluation and welfare comparisons. This paper reports estimates of the effects of JTPA training programs on the distribution of earnings. The estimation uses a new instrumental variable (IV) method that measures program impacts on the quantiles of outcome variables. This quantile treatment effects (QTE) estimator accommodates exogenous covariates and reduces to quantile regression when selection for treatment is exogenously determined. The QTE estimator can be computed as the solution to a convex linear programming problem, although this requires first-step estimation of a nuisance function. We develop distribution theory for the case where the first step is estimated nonparametrically. For women, the empirical results show that the JTPA program had the largest proportional impact at low quantiles. Perhaps surprisingly, however, JTPA training raised the quantiles of earnings for men only in the upper half of the trainee earnings distribution.

*We thank Moshe Buchinsky, Gary Chamberlain, Jinyong Hahn, Jerry Hausman, Whitney Newey, Shlomo Yitzhaki, and seminar participants at Berkeley, MIT-Harvard, Penn, and the Econometric Society Summer 1998 meetings for helpful comments and discussions. Thanks also go to Erik Beecroft at Abt Associates for providing us with the National JTPA Study data and for helpful discussions. Abadie acknowledges financial support from the Bank of Spain. Imbens acknowledges financial support from the Sloan Foundation.

1. Introduction

Effects of economic variables on distributions of outcomes are of fundamental interest in many areas of empirical economic research. A leading example is the question of how government programs affect the distribution of participants' earnings, since the welfare analysis of public policies involves distributions of outcomes. Policy-makers often hope that subsidized training programs will reduce earnings inequality by raising the lower quantiles of the earnings distribution and thereby reducing poverty (Lalonde (1995), US Department of Labor (1995)). Another example from labor economics is the effect of union status on the distribution of earnings. One of the earliest studies of the distributional consequences of unionism is Freeman (1980), while more recent analyses include Card (1996), and DiNardo, Fortin, and Lemieux (1996), who have asked whether changes in union status can account for a significant fraction of increasing wage inequality in the 1980s.

Although the importance of distribution effects is widely acknowledged, most evaluation research focuses on average outcomes, probably because the statistical techniques required to estimate effects on means are easier to use. Many econometric models also implicitly restrict treatment effects to operate in the form of a simple "location shift", in which case the mean effect captures the impact of treatment at all quantiles. Of course, the impact of treatment on a distribution is easy to assess when treatment status is randomly assigned and there is perfect compliance with treatment assignment. Randomization guarantees that outcomes in the treatment group are directly comparable to outcomes in the control group, so valid causal inferences can be obtained by simply comparing the treatment and control distributions. The problem of how to draw inferences about distributional effects in randomized studies with non-compliance or in observational studies with non-random assignment is more difficult, however, and has received less attention.[1]

In this paper, we show how to use a source of exogenous variation in treatment status

[1]Discussions of average treatment effects include Rubin (1977), Rosenbaum and Rubin (1983), and Heckman and Robb (1985). Manski (1994), Heckman, Smith and Clements (1997), Imbens and Rubin (1997), and Abadie (1999a) discuss effects on distributions. Manski (1994, 1997) develops estimators for bounds on quantiles.

– an instrumental variable – to estimate the effect of treatment on the quantiles of the distribution of outcomes in non-randomized studies, or in situations where the offer of treatment is randomized but treatment itself is voluntary. This Quantile Treatment Effects (QTE) estimator is used here to estimate the effect of training on trainees served by the Job Training Partnership Act (JTPA) of 1982, a large publicly-funded training program designed to help economically disadvantaged individuals. The data come from the National JTPA Study, a social experiment begun in the late early 1980s at 16 locations across the US to evaluate the effects of JTPA training. For this study, JTPA applicants were randomly assigned to treatment and control groups. Individuals in the treatment group were offered JTPA training, while those in the control group were excluded for a period of 18 months. Only 60 percent of the treatment group actually received training, but we can use the treatment assignment as an instrument for treatment.

The treatment effects estimated using the framework developed here are valid for a subpopulation we call *compliers*. This terminology is used because in randomized trials with partial compliance, like the JTPA, the relevant subpopulation consists of people who always comply with the treatment protocol. In fact, in the case of the JTPA, where (almost) no one in the control group received treatment, effects for compliers are also representative of effects on the treated.[2] In other cases, compliers are those whose treatment status is affected by an instrumental variable.

The identification results underlying the compliers approach to instrumental variables (IV) models were first established by Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). Imbens and Rubin (1997) extended these results to the identification of the effect of treatment on distributions, and Abadie (1999a) showed how to test global hypotheses about distribution impacts such as stochastic dominance. But neither of these papers developed simple estimators or a scheme for estimating the effect of treatment on quantiles. We focus here on conditional quantiles because quantiles provide useful summary

[2] Angrist and Imbens (1991) discuss the relationship between instrumental variables and effects on the treated. Orr, *et al.* (1996) and Heckman, Smith, and Taber (1994) report average effects on the treated in the JTPA. Heckman, Clements and Smith (1997) estimate the distribution of JTPA treatment effects using a non-IV framework.

statistics for distributions, and because quantile comparisons have been at the heart of recent discussions of changing wage inequality (see, e.g., Chamberlain (1991), Katz and Murphy (1992) and Buchinsky (1994)).

The paper is organized as follows. Section 2 outlines the conceptual framework and discusses the identification problem. Section 3 presents the estimator, which allows for a binary endogenous regressor (indicating exposure to treatment) and reduces to Koenker and Bassett (1978) quantile regression when selection for treatment is exogenous. Like quantile regression, the estimator developed here can be written as the solution to a convex linear programming (LP) problem, although implementation of the QTE estimator requires estimation of a nuisance function in a first step. Finally, Section 4 discusses the estimates of effects of training on the quantiles of trainee earnings. The estimates for women show larger proportional increases in earnings at lower quantiles of the trainee earnings distribution. But the estimates for men suggest the impact of training was largest in the upper half of the distribution and not at lower quantiles as policy-makers perhaps would have wished.

2. Conceptual Framework

The setup is as follows. The data consist of n observations on a continuously distributed outcome variable, Y, a binary treatment indicator D, and a binary instrument, Z. In the case of subsidized training, Y is earnings, D indicates program participation, and Z is an indicator of the randomized offer of training. Z and D are not equal in the JTPA because not everyone who was offered training received it and because a few people who were not offered training received services anyway. In a study of the effect of unions, Y might be a measure of wages, D would indicate union status, and Z would be an instrument for union status, say a dummy indicating individuals who work in firms that were subject to union organizing campaigns (Lalonde, Marschke and Troske (1996)). We also allow for an $r \times 1$ vector of covariates, X.

As in Rubin (1974, 1977) and our earlier work on instrumental variables estimation of causal effects, we define the causal effects of interest using potential outcomes and

potential treatment status. In particular, we define potential outcomes indexed against D, Y_d, and potential treatment status indexed against Z, D_z. Potential outcomes and potential treatment status describe possibly counterfactual states of the world. Thus, D_1 tells us what value D would take if Z were equal to 1, while D_0 tells us what value D would take if Z were equal to 0. Similarly, Y_d tells us what someone's outcome would be if they had $D = d$. The objects of causal inference are features of the distribution of potential outcomes, possibly restricted to particular subpopulations.

The observed treatment status is:

$$D = D_0 + (D_1 - D_0) \cdot Z.$$

In other words, if $Z = 1$, then D_1 is observed, while if $Z = 0$, then D_0 is observed. Likewise, the observed outcome variable is:

$$Y = Y_0 \cdot D + Y_1 \cdot (1 - D). \tag{1}$$

The reason why causal inference is difficult is that although we think of all possible counterfactual outcomes as being defined for everyone, only one potential treatment status and one potential outcome are ever observed for any one person.[3]

2.1. Principal Assumptions

The principal assumptions of the potential outcomes framework for IV are stated below:

Assumption 2.1: *For almost all values of X,*

(i) Independence: (Y_1, Y_0, D_1, D_0) *is jointly independent of Z given X.*

(ii) Non-Trivial Assignment: $P(Z = 1|X) \in (0, 1)$.

(iii) First-Stage: $E[D_1|X] \neq E[D_0|X]$.

(iv) Monotonicity: $P(D_1 \geq D_0|X) = 1$.

[3] The idea of potential outcomes appears in labor economics in discussions of the effects of union status. See, for example, Lewis' (1986) survey of research on union relative wage effects.

Assumption 2.1(i) subsumes two related requirements. First, comparisons by instrument status identify the causal effect of the instrument. This is equivalent to instrument-error independence in traditional simultaneous equations models. Second, potential outcomes are not directly affected by the instrument. This is an exclusion restriction. See Angrist, Imbens and Rubin (1996) for additional discussion of these two requirements and how they differ. Assumption 2.1(i) is plausible (though not guaranteed) in the case of the JTPA because of the randomly assigned offer of treatment.

Assumption 2.1(ii) requires that the conditional distribution of the instrument not be degenerate. The relationship between instruments and treatment assignment is restricted in two other ways as well. As in simultaneous equations models, we require that there be some correlation between D and Z; this is stated in Assumption 2.1(iii). Also, Imbens and Angrist (1994) have shown that Assumption 2.1(iv) guarantees identification of a meaningful average treatment effect in any model with heterogeneous potential outcomes that satisfies assumptions 2.1(i)-2.1(iii). This monotonicity assumption means that the instrument can only affect D in one direction. Monotonicity is plausible in most applications and it is automatically satisfied by latent-index models for treatment assignment.[4] It is also a reasonable assumption for the JTPA, where $D_0 = 0$ for (almost) everyone.

The inference problem in evaluation research involves comparisons of observed and counterfactual outcomes, possibly after conditioning on observed covariates, X. For example, many evaluation studies focus on estimating the difference between the average outcome for the treated (which is observed) and what this average would have been in the absence of treatment (which is counter-factual). Outside of a randomized trial, the difference in

[4] A latent-index model for participation is

$$D = 1\{\lambda_0 + Z \cdot \lambda_1 - \eta > 0\}$$

where λ_0 and λ_1 are parameters and η is an error term that is independent of Z. Then $D_0 = 1\{\lambda_0 > \eta\}$, $D_1 = 1\{\lambda_0 + \lambda_1 > \eta\}$, and either $D_1 \geq D_0$ or $D_0 \geq D_1$ for everyone. If $\lambda_1 < 0$ so that $D_0 \geq D_1$ for everyone, then monotonicity holds for $Z' = 1 - Z$.

average outcomes by observed treatment status is typically a biased estimate of this effect:

$$E[Y_1|X,D=1]-E[Y_0|X,D=0]=\{E[Y_1|X,D=1]-E[Y_0|X,D=1]\}$$
$$+\{E[Y_0|X,D=1]-E[Y_0|X,D=0]\}.$$

The first term in brackets is the average effect of the treatment on the treated, which can also be written as $E[Y_1-Y_0|X,D=1]$ since expectation is a linear operator; the second is the bias term. For example, comparisons of earnings by training status are biased if trainees are selected for training on the basis of low earnings potential. This bias extends to comparisons other than the mean. For example, the relationship above holds if we replace conditional expectations with conditional quantiles.

2.2. Identification Using Instrumental Variables

An instrumental variable solves the problem of identifying causal effects for a group of individuals whose treatment status is affected by the instrument. The following result (Imbens and Angrist (1994)) captures this idea formally:

LEMMA 2.1: *Under Assumption 2.1 (and assuming that the relevant expectations are finite)*

$$\frac{E[Y|X,Z=1]-E[Y|X,Z=0]}{E[D|X,Z=1]-E[D|X,Z=0]}=E[Y_1-Y_0|X,D_1>D_0].$$

$E[Y_1-Y_0|X,D_1>D_0]$ is called a Local Average Treatment Effect (LATE). We refer to individuals for whom $D_1>D_0$ as *compliers* because in a randomized trial with partial compliance, this group would consist of individuals who comply with the treatment protocol whatever their assignment. In other words, the set of compliers is the set of individuals whose treatment status was changed in the experiment induced by Z. Note that individuals in this set cannot be identified (i.e., we cannot name the people who are compliers) because we never observe both D_1 and D_0 for any one person. Also note that in the special case where $D_0=0$ for everyone,

$$\begin{aligned}E[Y_1-Y_0|X,D_1>D_0] &= E[Y_1-Y_0|X,D_1=1]=E[Y_1-Y_0|X,D_1=1,Z=1]\\ &= E[Y_1-Y_0|X,D=1],\end{aligned}$$

so LATE is the effect of treatment on the treated. The equivalence between effects for compliers and effects on the treated in cases where D_0 is identically zero holds for any distributional characteristic and not just means.

The compliers concept is at the heart of the LATE framework and provides a simple explanation for how IV methods work. Suppose initially that we could know who the compliers are. For these people, $Z = D$, since it is always true that $D_1 > D_0$. This observation plus Assumption 2.1 leads to the following lemma:

LEMMA 2.2: *Given Assumption 2.1 and conditional on X, treatment status, D, is ignorable (independent of the potential outcomes) for compliers:* $(Y_1, Y_0) \perp\!\!\!\perp D|X, D_1 > D_0$.

PROOF: Assumptions 2.1(i) says that $(Y_1, Y_0, D_1, D_0) \perp\!\!\!\perp Z|X$, so $(Y_1, Y_0) \perp\!\!\!\perp Z|X, D_1 = 1, D_0 = 0$. When $D_1 = 1$ and $D_0 = 0$, D can be substituted for Z. □

A consequence of Lemma 2.2 is that in the subpopulation of compliers, comparisons of means by treatment status estimate an average treatment effect even though treatment assignment is not ignorable in the population:

$$E[Y|X, D = 1, D_1 > D_0] - E[Y|X, D = 0, D_1 > D_0] = E[Y_1 - Y_0|X, D_1 > D_0]. \tag{2}$$

Of course, as it stands, Lemma 2.2 is of no practical use because the subpopulation of compliers is not identified (i.e., we do not observe D_1 and D_0 for the same individual). To make Lemma 2.2 operational, we define the following function of D, Z and X:

$$\kappa = 1 - \frac{D \cdot (1 - Z)}{1 - \pi_0(X)} - \frac{(1 - D) \cdot Z}{\pi_0(X)}, \tag{3}$$

where $\pi_0(X) = P(Z = 1|X)$. Note that κ equals one when $D = Z$, otherwise κ is negative. This function is useful because it "identifies compliers" in the following average sense:

LEMMA 2.3: *(Abadie, 1999b) Let $h(Y, D, X)$ be any integrable real function of (Y, D, X). Then, given Assumption 2.1,*

$$E[h(Y, D, X)|D_1 > D_0] = \frac{1}{P(D_1 > D_0)} \cdot E[\kappa \cdot h(Y, D, X)].$$

To see why this is true, note that, by monotonicity, the population can be partitioned into three groups: *compliers* who have $D_1 > D_0$, *always-takers* who have $D_1 = D_0 = 1$, and *never-takers* who have $D_1 = D_0 = 0$. Thus,

$$\begin{aligned} E[h(Y,D,X)|X,D_1>D_0] &= \frac{1}{P(D_1>D_0|X)}\Big\{E[h(Y,D,X)|X] \\ &- E[h(Y,D,X)|X,D_1=D_0=1]\cdot P(D_1=D_0=1|X) \\ &- E[h(Y,D,X)|X,D_1=D_0=0]\cdot P(D_1=D_0=0|X)\Big\}. \end{aligned}$$

Monotonicity means that all individuals with $Z = 1$ and $D = 0$ must be never-takers. Likewise, those with $Z = 0$ and $D = 1$ must be always-takers. Since Z is ignorable given X, we have the following expressions for always-takers and never-takers as a function of observed moments:

$$\begin{aligned} E[h(Y,D,X)|X,D_1=D_0=1] &= E[h(Y,D,X)|X,D=1,Z=0] \\ &= \frac{1}{P(D=1|X,Z=0)}\cdot E\left[\frac{D\cdot(1-Z)}{1-\pi_0(X)}\cdot h(Y,D,X)\middle| X\right], \end{aligned}$$

$$\begin{aligned} E[h(Y,D,X)|X,D_1=D_0=0] &= E[h(Y,D,X)|X,D=0,Z=1] \\ &= \frac{1}{P(D=0|X,Z=1)}\cdot E\left[\frac{(1-D)\cdot Z}{\pi_0(X)}\cdot h(Y,D,X)\middle| X\right]. \end{aligned}$$

Monotonicity and ignorability of Z given X can similarly be used to identify the proportions of always-takers and never-takers using $P(D_1 = D_0 = 1|X) = P(D = 1|X, Z = 0)$ and $P(D_1 = D_0 = 0|X) = P(D = 0|X, Z = 1)$. Integrating over X completes the argument.

An implication of Lemma 2.3 is that any parameter defined as the solution to a moment condition involving (Y, D, X) is identified for compliers. This point is explored in detail in Abadie (1999b).[5] In the next section, we show how Lemma 2.3 can be used to develop an

[5] For example, if we define μ and α as

$$(\mu,\alpha) = \text{argmin}_{(m,a)}\, E[(Y-m-aD)^2|D_1>D_0],$$

then, $\mu = E[Y_0|D_1 > D_0]$, and $\alpha = E[Y_1 - Y_0|D_1 > D_0]$, so that α is LATE (although μ is not the same intercept that is identified by conventional IV methods). By Lemma 2.3, (μ, α) also minimizes $E[\kappa \cdot (Y - m - aD)^2]$.

estimator for the causal effect of treatment on the quantiles of an outcome variable.

3. Quantile Treatment Effects

3.1. The QTE Model

The QTE estimator is based on a model where the effect of treatment and covariates is linear and additive at each quantile, so that a single treatment effect is estimated. The analysis is straightforward when the treatment effect varies with X, but we use an additive model because the resulting estimator simplifies to Koenker and Bassett (1978) quantile regression when there is no instrumenting. The relationship between QTE and quantile regression is therefore analogous to the relationship between conventional IV and ordinary least squares (OLS).

The parameters of interest are defined as follows:

Assumption 3.1: *For $\theta \in (0,1)$, there exist unique $\alpha_\theta \in \mathbb{R}$ and $\beta_\theta \in \mathbb{R}^r$ such that*

$$Q_\theta(Y|X, D, D_1 > D_0) = \alpha_\theta D + X'\beta_\theta. \tag{4}$$

where $Q_\theta(Y|X, D, D_1 > D_0)$ denotes the θ-quantile of Y given X and D for compliers.

As a consequence of Lemma 2.2, the parameter of primary interest in this model, α_θ, gives the difference in the θ-quantiles of Y_1 and Y_0 for compliers. This tells us, for example, whether JTPA training changed the median earnings of participants. Note, however, that in contrast with average treatment effects, where average differences equal differences in averages, α_θ is not the quantile of the difference $(Y_1 - Y_0)$. Although the latter may also be of interest, we focus on the marginal distributions of potential outcomes because identification of the distribution of $Y_1 - Y_0$ requires much stronger assumptions and because economists making social welfare comparisons typically use differences in distributions and not the distribution of differences for this purpose (see, e.g., Atkinson (1970)).[6]

[6]Heckman, Smith and Clements (1997) discuss models where features of the distribution of the difference $(Y_1 - Y_0)$ are identified. They note that this may be of interest for questions regarding the political economy of social programs. If the ranking of individuals in the distribution of the outcome is preserved

The model above differs in a number of ways from the model in the seminal papers by Amemiya (1982) and Powell (1983), who used least absolute deviations to estimate a simultaneous equations system. Their approach begins with a traditional simultaneous equations model, and is not motivated by an attempt to characterize effects on distributions. Rather, the idea is to improve on 2SLS when the distributions of the error terms are long-tailed. Most importantly, in contrast with the parameters in equation (4), the parameters of interest in the Amemiya/Powell setup do not, in general, define a conditional quantile function.[7]

The parameters of the conditional quantile function in equation (4) can be expressed as (see Bassett and Koenker (1982)):

$$(\alpha_\theta, \beta_\theta) = \text{argmin}_{(\alpha,\beta)\in\mathbb{R}^{r+1}}\ E[\rho_\theta(Y - \alpha D - X'\beta)|D_1 > D_0],$$

where $\rho_\theta(\lambda)$ is the check function, defined as $\rho_\theta(\lambda) = (\theta - 1\{\lambda < 0\}) \cdot \lambda$ for any real λ. Therefore, using Lemma 2.3, α_θ and β_θ are identified as

$$(\alpha_\theta, \beta_\theta) = \text{argmin}_{(\alpha,\beta)\in\mathbb{R}^{r+1}}\ E[\kappa \cdot \rho_\theta(Y - \alpha D - X'\beta)]. \tag{5}$$

This population objective function is globally convex in $(\alpha_\theta, \beta_\theta)$ since it is equal to the check-function minimand for compliers times some positive constant $(P(D_1 > D_0))$. Following the analogy principle (Manski (1988)) a natural estimator of $(\alpha_\theta, \beta_\theta)$ is the sample counterpart of (5). However, since κ is negative when D is not equal to Z, the sample objective function turns out to be non-convex. A number of algorithms exist for minimization problems of this type (piecewise linear and non-convex objective functions), but they do not ensure a global optimum (see, e.g., Charnes and Cooper (1957) or Fitzenberger (1997a,b), for a discussion of a related censored quantile regression problem). Unlike the

under the treatment, then the estimator in this paper is informative about the distribution of treatment impacts. King (1983) discusses horizontal equity concerns that require welfare analyses involving the joint distribution of outcomes.

[7]Identification in the Amemiya and Powell papers comes from conditional median restrictions on the reduced form. However, a conditional median restriction on the reduced form does not imply that the structural equation is a conditional median. In fact, for a binary endogenous regressor, conditional median restrictions on the reduced form and structural equation are typically incompatible.

conventional quantile regression minimand, the sample analog of equation (5) does not have a linear programming representation.

Now, let $U = (Y, D, X)$; applying the Law of Iterated Expectations to equation (5), we obtain

$$(\alpha_\theta, \beta_\theta) = \text{argmin}_{(\alpha,\beta)\in^{r+1}} E[\kappa_\nu \cdot \rho_\theta(Y - \alpha D - X'\beta)], \tag{6}$$

where

$$\kappa_\nu = E[\kappa|U] = 1 - \frac{D \cdot (1 - \nu_0(U))}{1 - \pi_0(X)} - \frac{(1 - D) \cdot \nu_0(U)}{\pi_0(X)}$$

for $\nu_0(U) = E[Z|U] = P(Z = 1|Y, D, X)$. Although simple to derive, this second representation is of signal importance because, as we show below, κ_ν is a conditional probability and is therefore non-negative.

LEMMA 3.1: *Under Assumption 2.1,* $\kappa_\nu(U) = P(D_1 > D_0|U)$.

PROOF: First consider the product $D \cdot (1 - Z)$. This differs from zero only if $Z = 0$ and $D_0 = 1$. By monotonicity, $D_0 = 1$ implies $D_1 = 1$. Hence:

$$\begin{aligned} E[D \cdot (1 - Z)|U] &= P(D(1 - Z) = 1|U) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|D_1 = D_0 = 1, U) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|D_1 = D_0 = 1, Y_1, X) \\ &= P(D_1 = D_0 = 1|U) \cdot P(Z = 0|X). \end{aligned}$$

Similarly, $E[(1 - D) \cdot Z|U] = P(D_1 = D_0 = 0|U) \cdot P(Z = 1|X)$. Therefore,

$$\begin{aligned} \kappa_\nu(U) &= E\left[1 - \frac{D(1 - Z)}{P(Z = 0|X)} - \frac{(1 - D)Z}{P(Z = 1|X)} \middle| U\right] \\ &= 1 - P(D_1 = D_0 = 1|U) - P(D_1 = D_0 = 0|U) = P(D_1 > D_0|U). \end{aligned}$$

□

A consequence of this lemma is that it is possible to develop a QTE estimator with an LP representation based on a sample analog of equation (6)). This can be thought of as

a scheme to "convexify" the sample analog of (5). The resulting convex QTE estimator minimizes a positively-weighted check-function minimand, with a global minimum that can be obtained as the solution to a linear programming problem in a finite number of simplex iterations. This is similar in spirit to Buchinsky and Hahn's (1998) LP-type estimator for censored quantile regression.

An interesting question is whether there is any efficiency cost to using an estimator based on κ_ν instead of the sample analog of (5). In fact, it can be shown that both strategies produce estimators that are asymptotically equivalent.[8] In light of this result, the rest of the paper focuses on estimation by minimizing the sample analog of (6). This requires first-step estimation of $\pi_0(X)$ and $\nu_0(U)$ to construct an estimate of $\kappa_\nu(U)$, denoted $\widehat{\kappa}_\nu$. The distribution theory is developed assuming that X is discrete, so a saturated linear model consistently estimates $\pi_0(X)$. We use series approximation to estimate $\nu_0(U)$ non-parametrically in (X, D) cells.[9] If the number of terms in the series approximation increases at an appropriate rate with the sample size, this procedure ensures that the estimated conditional expectations converge to the true conditional expectations. In practice, however, it may make sense to use something less than a saturated model for X when the dimensionality of X is high. We discuss practical aspects of the estimation strategy further when the results are presented in Section 4.

3.2. Estimation

Assume that we have a random sample $\{Y_i, D_i, X_i, Z_i\}_{i=1}^n$. Let $W = (D, X')'$ and $\delta_\theta = (\alpha_\theta, \beta_\theta')'$ for α_θ and β_θ in equation (4). If κ_ν were known, the estimation problem would reduce to a weighted quantile regression problem of the type discussed by Newey and Powell (1990). Since κ_ν is unknown, we estimate this function nonparametrically in a first step

[8] See Newey (1994); the estimators are asymptotically equivalent since they nonparametrically estimate the same functional.

[9] Buchinsky and Hahn (1998) similarly decompose the covariates in a censored quantile regression problem into a set of discrete variables and a set of continuous variables. They use non-parametric methods to estimate the conditional probability of censoring in cells defined by the discrete variables.

and use the fitted values $\widehat{\kappa}_\nu(U_i)$ in a second step to construct the estimator:

$$\widehat{\delta}_\theta = \operatorname{argmin}_{\delta \in \mathbb{R}^{r+1}} \frac{1}{n} \sum_{i=1}^{n} 1\{\widehat{\kappa}_\nu(U_i) \geq 0\} \cdot \widehat{\kappa}_\nu(U_i) \cdot \rho_\theta(Y_i - W_i'\delta), \tag{7}$$

First step estimation of κ_ν is carried out using non-parametric series regression. For an increasing sequence of positive integers $\{\lambda(k)\}_{k=1}^{\infty}$ and a positive integer K, let $p^K(Y) = (Y^{\lambda(1)}, ..., Y^{\lambda(K)})$. Assume that X only takes on a finite number of values (so that $W \in \{w_1, ..., w_J\}$). Then, any random sample $\{V_i\}_{i=1}^{n} = \{(Z_i, U_i)\}_{i=1}^{n}$ from $V = (Z, U)$ can be indexed as $\{\{V_{i_j}\}_{i_j=1}^{n_j}\}_{j=1}^{J}$, where $\{V_{i_j}\}_{i_j=1}^{n_j}$ are subsequences for distinct fixed values of (X, D). In the same fashion, the sample can be indexed as $\{\{V_{i_l}\}_{i_l=1}^{n_l}\}_{l=1}^{L}$, where $\{V_{i_l}\}_{i_l=1}^{n_l}$ are subsequences for distinct fixed values of X. Now, a nonparametric power series estimator $\widehat{\nu}(U)$ of $\nu_0(U)$ is given by the Least Squares projection of $\{Z_{i_j}\}_{i_j=1}^{n_j}$ on $\{p^K(Y_{i_j})\}_{i_j=1}^{n_j}$ (this amounts to non-parametric series regression of Z on Y in each W-cell). Let $\widehat{\nu}_i$ be the fitted values of such estimator for the observations in our sample. Consider the simple estimator $\widehat{\pi}(X)$ of $\pi_0(X)$ obtained by averaging Z within cells of X. Our first step estimator of κ_ν is given by:

$$\widehat{\kappa}_\nu(U_i) = 1 - \frac{D_i \cdot (1 - \widehat{\nu}_i)}{1 - \widehat{\pi}(X_i)} - \frac{(1 - D_i) \cdot \widehat{\nu}_i}{\widehat{\pi}(X_i)}.$$

3.3. Distribution Theory

This subsection summarizes asymptotic results for the QTE estimator. Proofs are given in the appendix.

Theorem 3.1: *Under assumptions 2.1 and 3.1 and if (i) the data are i.i.d.; (ii) conditional on W, Y is continuously distributed with support equal to a compact interval and density bounded away from zero; (iii) $\pi_0(X)$ is bounded away from zero and one, and X takes on a finite number of values; (iv) conditional on W, ϵ_θ is continuously distributed with bounded density; the distribution function of ϵ_θ conditional on W and $D_1 > D_0$ is continuously differentiable at zero with density $f_{\epsilon_\theta|W,D_1>D_0}(0)$ that is bounded and bounded away from zero uniformly in W; (v) κ_ν is bounded away from zero uniformly in U; (vi) for s equal to the*

number of continuous derivatives in Y *of* ν_0, $n \cdot K^{-2s} \to 0$ *and* $K^5/n \to 0$. *Then,* $n^{1/2}(\widehat{\delta}_\theta - \delta_\theta) \xrightarrow{d} \mathcal{N}(0, \Omega)$, *where* $\Omega = J^{-1}\Sigma J^{-1}$, $J = E[f_{\epsilon_\theta|W,D_1>D_0}(0) \cdot WW'|D_1 > D_0] \cdot P(D_1 > D_0)$ *and* $\Sigma = E[\psi\psi']$ *with* $\psi = \kappa \cdot m(U) + H(X) \cdot \{Z - \pi_0(X)\}$, $m(U) = (\theta - 1\{Y - W'\delta_\theta < 0\}) \cdot W$,

$$H(X) = E\left[m(U) \cdot \left(-\frac{D \cdot (1-Z)}{(1-\pi_0(X))^2} + \frac{(1-D) \cdot Z}{(\pi_0(X))^2} \right) \middle| X \right].$$

The asymptotic variance formula provided by this theorem is robust to mis-specification of the functional form (in Assumption 3.1). In such a case, quantile regression estimates the best linear predictor under asymmetric loss.[10]

To produce an estimator of the asymptotic variance matrix, let

$$\varphi_h(\delta) = \frac{1}{h}\varphi\left(\frac{Y - W'\delta}{h}\right) \qquad \text{and} \qquad \varphi_{h,i}(\delta) = \frac{1}{h}\varphi\left(\frac{Y_i - W_i'\delta}{h}\right)$$

where $\varphi(\cdot)$ is a kernel function. Consider the following estimator of J:

$$\widehat{J} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\kappa}_\nu(U_i) \cdot \varphi_{h,i}(\widehat{\delta}_\theta) \cdot W_i W_i'.$$

For i in the l-cell of X, let

$$\widehat{H}_i = \frac{1}{n_l}\sum_{i_l=1}^{n_l} (\theta - 1\{Y_{i_l} - W_{i_l}'\widehat{\delta}_\theta < 0\}) \cdot W_{i_l} \cdot \left(\frac{(1-D_{i_l}) \cdot Z_{i_l}}{(\widehat{\pi}(X_i))^2} - \frac{D_{i_l} \cdot (1-Z_{i_l})}{(1-\widehat{\pi}(X_i))^2} \right),$$

$$\widehat{\kappa}(V_i) = 1 - \frac{D_i \cdot (1-Z_i)}{1-\widehat{\pi}(X_i)} - \frac{(1-D_i) \cdot Z_i}{\widehat{\pi}(X_i)},$$

$$\widehat{\psi}_i = \widehat{\kappa}(V_i) \cdot (\theta - 1\{Y_i - W_i'\widehat{\delta}_\theta < 0\}) \cdot W_i + \widehat{H}_i \cdot \{Z_i - \widehat{\pi}(X_i)\}.$$

An estimator of Σ can then be constructed as

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\psi}_i \widehat{\psi}_i{}'.$$

The following theorem establishes the consistency of an asymptotic covariance matrix estimator.

[10]Most of the literature on quantile regression treats the linear model as a literal specification for conditional quantiles. Alternately, the linear model can be viewed as an approximation. This interpretation is discussed by Buchinsky (1991), Chamberlain (1991), Fitzenberger (1997), and Portnoy (1991).

THEOREM 3.2: *Under the assumptions of Theorem 3.1 and (i)* $h \to 0$, $nh^4 \to \infty$; *(ii) for some neighborhood of* 0, $f_{\epsilon_\theta|W,D_1>D_0}(\cdot)$ *has bounded and continuous first derivative; (iii)* $\varphi(z) \geq 0$, $\int \varphi(z)\,dz = 1$, $\int |z \cdot \varphi(z)|\,dz < \infty$; *(iv) there exists* $C > 0$ *such that* $|\varphi(z) - \varphi(z_0)| \leq C \cdot |z - z_0|$. *Then* $\widehat{\Omega} = \widehat{J}^{-1}\,\widehat{\Sigma}\,\widehat{J}^{-1} \xrightarrow{p} \Omega$.

4. EFFECTS OF SUBSIDIZED TRAINING

4.1. BACKGROUND

The JTPA began funding training in October 1983, and continued to fund federally-sponsored training programs into the late 1990s. The program included a number of parts or "titles", the largest of which is Title II, which supports training for those judged to be economically disadvantaged. At the time of the National JTPA Study in the early 1990s, JTPA Title II programs were serving about 1 million participants a year, at an annual cost of roughly 1.6 billion dollars. JTPA services were delivered at 649 sites, also called Service Delivery Areas (SDAs), located throughout the country.

Title II of the JTPA is unusual in that it explicitly incorporated a mandate for randomized evaluation.[11] The National JTPA study is the largest randomized training evaluation ever undertaken in the US. The JTPA evaluation study collected data on about 20,000 participants at 16 SDAs. These sites were not a random sample of all SDAs; rather, they were chosen for diversity, willingness and ability to implement the experimental design, and the size and composition of the experimental sample they could provide. Although the non-random selection of sites raises issues of external validity (as in many clinical trials), within sites, applicants were randomly selected for JTPA treatment. The evaluation sample includes applicants who applied between November 1987 and September 1989.

The original study of the labor-market impact of Title II services was based on 15,981 persons for whom continuous data on earnings (from either State unemployment insurance (UI) records or two follow-up surveys) were available for at least 30 months after random

[11] Other parts of the JTPA, such as Title III programs for workers who lost their jobs as a consequence of international competition, did not include a randomized evaluation. Background for this section is drawn from Orr, *et al.* (1996), Bloom, *et al.* (1997), and the US Department of Labor (1999) website.

assignment. Although data are available on a range of labor market outcomes for this sample, we focus on the sum of earnings in this 30-month period since this is probably the best measure of the program's lasting economic impact on participants. Individuals who were not offered treatment were generally excluded from receiving JTPA services for period of 18 months following their application (though they could participate in other programs at any time).

The JTPA was a complicated program that offered a wide range of services. JTPA service providers included community colleges, State employment services, community organizations, and private-sector training agencies. The types of services offered can be grouped into three general service strategies. These strategies are (i) classroom training in occupational skills, basic education, or both; (ii) on-the-job training and/or job search assistance (OJT/JSA); (iii) other services that may have included probationary employment and/or a combination of the first two. For the National JTPA Study, service strategies were recommended as part of the JTPA intake process, before random assignment. Although individuals were assigned to treatment with different probabilities depending on their SDA, the data in the analysis sample were artificially balanced to maintain a 2/1 treatment-control ratio at each location.

The JTPA offered services to a number of different groups. Title II applicants were generally deemed eligible for training if they faced one of a number of "barriers to employment". These included long-term use of welfare, being a high school dropout, 15 or more recent weeks of unemployment, limited English proficiency, physical or mental disability, reading proficiency below 7th grade level, or an arrest record. The most common barriers were unemployment spells and high-school dropout status. Applicants were categorized as being in one of five groups: adult men, adult women, female youth, male youth non-arrestees, and male youth arrestees. In this study we focus on adult men and women because the samples are largest for these two groups. There are 6,102 adult women with 30-month earnings data and 5,102 adult men with 30-month earnings data.

4.2. Average Effects

Using our earlier notation, Y is 30-month earnings, D indicates those who were recorded as having been enrolled for JTPA services, and Z indicates the offer of services. Although the offer of treatment was randomly assigned, only about 60 percent of those offered training actually received JTPA services. This is a consequence of the JTPA evaluation design, which randomized the offer of services early in the application process, but did not compel those offered services to participate in training.

While all applicants indicated at least some interest in receiving JTPA services, those offered treatment were not necessarily notified immediately, and some time may have passed before training could begin. In the meantime, applicants selected for treatment may have found jobs, received services somewhere else, or simply lost interest. Providers may also have had an incentive to delay the enrollment of applicants that they thought were unlikely to benefit from treatment, while some SDAs were unable to find service providers for some applicants. Also, on the other side of the randomization offer, a small proportion of those selected for the control group (1.6 percent) received JTPA services despite the experimenters' attempt to prevent this.[12] Treatment status is therefore likely to be correlated with potential outcomes and cannot be treated as exogenous.

Although treatment status itself was not randomly assigned, the assumptions of our framework appear to apply in this case: the randomized offer of treatment is likely to have been independent of potential outcomes, the offer of treatment is unlikely to have affected outcomes through any mechanism other than treatment itself, and denial of services by randomization is not likely to have made treatment more likely. Moreover, because of the very low probability of receiving JTPA services in the control ($Z = 0$) group, effects for compliers in this case can also be interpreted as effects on those who were treated.

Since training offers were randomized in the National JTPA Study, covariates (X) are not required to identify training effects. Even in experiments like this, however, it is

[12] Adult men and women who were offered treatment ultimately received about 150 more hours of training services than were received by the members of the control group.

customary to control for covariates to correct for chance associations between D and X (as in Orr, *et al.*, 1996). Moreover, in our setup, covariates can be used to describe earnings quantiles for compliers in population subgroups, since we estimate $Q_\theta(Y|X, D, D_1 > D_0)$. We therefore include as covariates dummies for black and Hispanic applicants, a dummy for high-school graduates (including GED holders), dummies for married applicants, 5 age-group dummies, and dummies for AFDC receipt (for women) and whether the applicant worked at least 12 weeks in the 12 months preceding random assignment. Also included are dummies for the original recommended service strategy (classroom, OJT/JSA, other) and a dummy for whether earnings data are from the second follow-up survey.[13] In addition, the analysis is carried out separately for men and women since previously reported results differed by sex.

Descriptive statistics are reported in Table I. There are more minority applicants than in the general population and, not surprisingly given the program rules, a relatively low proportion of high school graduates. The applicants also have low previous employment rates. Most of the men were recommended for OJT/JSA services, while the women were slightly more likely to be recommended for classroom training than OJT/JSA or other services. Average 30-month earnings in the sample are about \$19,000 for men and \$13,000 for women.

As a benchmark for the purposes of comparison with earlier analyses of the JTPA, Table II reports OLS and conventional instrumental variables (2SLS) estimates of the impact of training.[14] The first column reports unadjusted trainee/non-trainee differences, while the OLS estimates in column (2) are from a regression of the dependent variable on the covariates and a training dummy (D). Without the use of covariates, the training/non-training difference is \$3,970 for men and \$2,133 for women. Trainee/nontrainee differences are precisely measured for both men and women. OLS estimates of training effects in

[13]The covariate information comes from a background survey conducted as part of the JTPA intake process. The covariate list used here is similar to that described in Appendix B of Orr *et al.* (1996), except that we collapsed some categories and omitted SDA dummies because they had low explanatory power.

[14]As with the quantile regression and QTE estimates discussed later, the standard errors in Table II are robust in the sense that they provide consistent estimates of the asymptotic variance of the estimators under general misspecification.

models with covariates are similar to the differences without controls.

The reduced-form effects of the offer of treatment are reported in columns (3) and (4). Not surprisingly, since Z was randomly assigned without conditioning on covariates, the estimates with and without covariates differ little. Note that the reduced form estimates are not directly comparable with the OLS estimates since many of those offered training did not actually receive training.

The instrumental variable estimates in columns (5) and (6) of Table II use the randomized offer of treatment (Z) as an instrument for D in the same regression as was used to construct the estimates in columns (1) and (2). This corrects for non-participation among those offered training. When covariates are included, the 2SLS estimate for men is \$1,593 with a standard error of \$895, less than half the size of the corresponding OLS estimate. For women, however, the 2SLS estimate is \$1,780 with a standard error of \$532, not dramatically different from the corresponding OLS estimate. This amounts to a 9 percent earnings increase for men and a 15 percent earnings increase for women. These results are similar to those reported in previous studies.[15]

4.3. Estimates of Quantile Treatment Effects

Table III reports OLS and conventional quantile regression estimates of the effect of training. The covariates are the same as those used to construct the estimates in Table II. The OLS estimate of the training coefficient is \$3,754 for men and \$2,215 for women. The quantile regression estimates show that the gap in quantiles by trainee status is much larger (in proportionate terms) below the median than above it. For men, the .85 quantile of trainee earnings is about 13 percent higher than the corresponding quantile for non-trainees, while the .15 quantile is 136 percent higher. For women the difference in impact across quantiles is less dramatic, but still marked. Like the OLS estimates shown in the table, the quantile regression coefficients do not necessarily have a causal interpretation. Rather they provide

[15]The 2SLS estimates in Table II are very close to those in Table 4.6 of the National JTPA study by Orr *et al.* (1996). The estimates are not identical because the covariates are not identical. Percentage effects were computed as the coefficient on training, divided by fitted values with the training dummy set to zero and other covariates set to means for the treated.

a descriptive comparison of earnings distributions for trainees and non-trainees.

Implementation of the QTE estimator requires first step estimation of κ_ν. The theoretical results in the previous section are based on non-parametric series estimation of the conditional expectations in κ_ν, but this leaves open a range of possibilities. Since the elements of X are discrete, non-parametric estimation of $E[Z|X]$ is in principle straightforward. In practice, however, a fully saturated model leads to problems with small or missing covariate cells. We therefore estimated $E[Z|X]$ using the restriction that because of random assignment, Z and X should be uncorrelated in large samples. The resulting estimate is simply the empirical $E[Z]$. The expectation $\nu_0(U) = E[Z|Y, D, X]$ was estimated using separate models for $D = 0, 1$. Most X's were dropped because they had little explanatory value. A series approximation was used to estimate terms in Y. Selection of the order for the series approximation was guided by cross-validation. The order is the same for both values of D.[16]

QTE estimates of the effect of training on median earnings, reported in Table IV, are similar in magnitude though less precisely estimated than the 2SLS estimates in Table II. As noted earlier, for women the 2SLS estimates are not much smaller than OLS estimates, but for men the 2SLS estimates are considerably smaller than OLS.

A particularly interesting finding for men is that the QTE estimates of effects on quantiles exhibit a pattern very different from the quantile regression estimates. In particular, the QTE estimates show no evidence of a change in the .15 or .25 quantile. The estimates at low quantiles are substantially smaller than the corresponding quantile regression estimates, and they are small in absolute terms. For example, the QTE estimate (standard error) of the effect on the .15 quantile for men is \$121 (475), while the corresponding quantile regression estimate is \$1,187 (205). Similarly, the QTE estimate (standard error) of the effect on the .25 quantile for men is \$702 (670), while the corresponding quantile regression

[16] Hausman and Newey (1995) use a similar approach to dimension-reduction for non-parametric estimation of consumer demand equations. Given estimates of κ_ν, we computed QTE coefficient estimates by weighted quantile regression using the Barrodale-Roberts (1973) linear programming algorithm for quantile regression (see, e.g., Koenker and D'Orey (1987)). A biweight kernel was used for the estimation of standard errors.

estimate is \$2,510 (356). This suggests that quantile regression estimates of training effects at low quantiles are especially distorted by positive selection on earnings potential. It seems that training did not really change the lower deciles of the trainee earnings distribution for men. In contrast with the results at low quantiles, however, the QTE estimates of effects on male earnings above the median are large and statistically significant (though still smaller than the corresponding quantile regression estimates).

The QTE estimates for women show significant effects of training at every quantile, with the largest proportional effects at low quantiles. For example, training is estimated to raise the .15 quantile of earnings for women by \$324 (175), an increase of 35 percent. The estimates also suggest training raises the .85 quantile by \$1,900 (997), but this is an increase of only 7 percent. Most of the QTE estimates for women are reasonably close to the corresponding quantile regression estimates. Thus, whether or not training is treated as endogenous, the estimates support the notion that for women training had a bigger proportional impact on the lower tail of the earnings distribution than the upper tail. Of course, women's earnings are especially low in this sample, so large proportional effects do not translate into large dollar amounts.[17]

Orr, *et al.* (1996) reported effects by subgroups but found no clear patterns. They concluded that (p. 160) "the benefits of JTPA are broadly distributed across a wide variety of different types of men and women." Heckman, Smith, and Clements (1997) similarly concluded that heterogeneity of impacts is important but that most women benefited from the JTPA. Our results do not contradict these general conclusions, but they nevertheless show more heterogeneity in program effects than is revealed by a simple analysis within subgroups. In particular, our results strongly suggest that training for adult women had a much larger proportional effect on the lower tail of the earnings distribution than on the upper tail (though the absolute effect on the lower tail is small).

[17]Interestingly, QTE estimates of the proportional effects of training on men are smaller than conventional quantile regression estimates not only because the training impact is lower, but also because the constant is bigger. This reflects the fact that the constant and covariate-effects estimated by QTE are for $Q_\theta(Y_0|X, D_1 > D_0)$. This is bigger than the quantile regression intercept because of positive selection for male compliers.

Perhaps most striking among our findings is the result that training for adult men does not seem to have raised the lower quantiles of their earnings. This may be because of an effort by program operators to target services at relatively easy-to-employ men with higher earnings potential. The results in distributional changes that would be undesirable in any assessment using a social welfare function that weights the lower tail of the earnings distribution more heavily. Since the ostensible purpose of the JTPA was to aid economically disadvantaged workers, it seems likely that the lower quantiles are of particular concern to policy makers. One response to this finding might be that few JTPA applicants were very well off, so that distributional effects within applicants are of less concern than the fact that the program helped many applicants overall. However, the upper quantiles of earnings were reasonably high for adult males who participated in the National JTPA Study. Increasing this upper tail is therefore unlikely to have been a high priority.

5. Summary and Conclusions

This paper reports estimates of the effect of subsidized training on the quantiles of earnings for participants. We use a new estimator for the effect of a non-ignorable treatment on quantiles. The QTE estimator can be used to determine how an intervention affects the distribution of any variable for individuals whose treatment status is changed by a binary instrument. The estimator accommodates exogenous covariates and collapses to conventional quantile regression when the treatment is exogenous. It minimizes a convex piecewise-linear objective function similar to that for conventional quantile regression, and can be computed as the solution to a linear programming problem after first-step estimation of a nuisance function. The paper develops distribution theory for the case where this first step is estimated nonparametrically. QTE estimates of the effect of training on the quantiles of the earnings distribution suggest interesting and important differences in program effects at different quantiles, and differences in distributional impact for men and women. These differences are large enough to potentially change the welfare analysis of the JTPA program.

Appendix

Proof of Theorem 3.1:

This proof largely follows that of Theorem 1 in Buchinski and Hahn (1998). Consider,

$$G_n(\tau,\kappa)=\sum_{i=1}^{n} g_i(\tau,\kappa)$$

where

$$g_i(\tau,\kappa)=\kappa(U_i)\cdot\{\theta\cdot[(\epsilon_{\theta i}-n^{-1/2}\,W_i'\tau)^+-\epsilon_{\theta i}^+]+(1-\theta)\cdot[(\epsilon_{\theta i}-n^{-1/2}\,W_i'\tau)^--\epsilon_{\theta i}^-]\},$$

and $\epsilon_{\theta i}=Y_i-W_i'\delta_\theta$. The function $G_n(\tau,1\{\widehat{\kappa}_\nu\geq 0\}\cdot\widehat{\kappa}_\nu)$ is convex in τ and it is minimized at $\tau_n=\sqrt{n}(\widehat{\delta}_\theta-\delta_\theta)$. Now, define $\Gamma_n(\tau,\kappa)=E[G_n(\tau,\kappa)]$. Note that,

$$\frac{\partial g_i(\tau,\kappa_\nu)}{\partial\tau}=-n^{-1/2}\,W_i\,\kappa_\nu(U_i)\cdot(\theta-1\{\epsilon_{\theta i}-n^{-1/2}\,W_i'\tau<0\})$$

almost surely. By (iv) and Weierstrass domination,

$$\frac{\partial E[g(\tau,\kappa_\nu)]}{\partial\tau}\,|_{\tau=0}=-n^{-1/2}\,E[W\,\kappa_\nu(U)\cdot(\theta-1\{\epsilon_\theta<0\})]=0,$$

$$\frac{\partial^2 E[g(\tau,\kappa_\nu)]}{\partial\tau\partial\tau'}\,|_{\tau=0}=n^{-1}\,E[f_{\epsilon_\theta|W,D_1>D_0}(0)\cdot WW'|D_1>D_0]\cdot P(D_1>D_0).$$

Then,

$$\Gamma_n(\tau,\kappa_\nu)=\frac{1}{2}\tau' J\tau+o(1),$$

where $J=E[f_{\epsilon_\theta|W,D_1>D_0}(0)\cdot WW'|D_1>D_0]\cdot P(D_1>D_0)$. Note that by (ii) and since both κ_ν and $f_{\epsilon_\theta|W,D_1>D_0}(0)$ are bounded away from zero, J is non-singular. Define,

$$\zeta_n(U_i)=n^{-1/2}(\theta-1\{\epsilon_{\theta i}<0\})\cdot W_i,$$

$$\omega_n(\kappa)=\sum_{i=1}^{n}\kappa(U_i)\cdot\zeta_n(U_i),$$

and

$$\rho_n(U_i,\kappa,\tau)=\kappa(U_i)\cdot\{\theta\cdot[(\epsilon_{\theta i}-n^{-1/2}\,W_i'\tau)^+-\epsilon_{\theta i}^+]+(1-\theta)\cdot[(\epsilon_{\theta i}-n^{-1/2}\,W_i'\tau)^--\epsilon_{\theta i}^-]+\tau'\zeta_n(U_i)\}.$$

Note that $E[\omega_n(\kappa_\nu)]=0$, then

$$\begin{aligned}G_n(\tau,\kappa)&=\Gamma_n(\tau,\kappa_\nu)+(G_n(\tau,\kappa)-\Gamma_n(\tau,\kappa_\nu))\\&=\Gamma_n(\tau,\kappa_\nu)-\tau'\omega_n(\kappa)+(G_n(\tau,\kappa)+\tau'\omega_n(\kappa)-\{\Gamma_n(\tau,\kappa_\nu)+\tau' E[\omega_n(\kappa_\nu)]\})\\&=\Gamma_n(\tau,\kappa_\nu)-\tau'\omega_n(\kappa)+\sum_{i=1}^{n}\{\rho_n(U_i,\kappa,\tau)-E[\rho_n(U,\kappa_\nu,\tau)]\}\end{aligned}$$

Lemma A.1 :

$$\omega_n(1\{\widehat{\kappa}_\nu\geq 0\}\cdot\widehat{\kappa}_\nu)=n^{-1/2}\sum_{i=1}^{n}\psi_i+o_p(1)\qquad\text{with}\qquad E\psi=0\qquad\text{and}\qquad E\|\psi\|^2<\infty.$$

LEMMA A.2 : $\sum_{i=1}^{n}\{\rho_n(U_i, 1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu, \tau) - E[\rho_n(U, \kappa_\nu, \tau)]\} = o_p(1)$.

Applying Lemma A.2:

$$G_n(\tau, 1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu) = \frac{1}{2}\tau' J\tau - \tau'\omega_n(1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu) + o_p(1),$$

for a given τ. Let $\eta_n = J^{-1}\omega_n(1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu)$. Note that:

$$\frac{1}{2}(\tau - \eta_n)' J(\tau - \eta_n) = \frac{1}{2}\tau' J\tau - \tau'\omega_n(1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu) + \frac{1}{2}\eta_n' J\eta_n.$$

Define $\lambda_n(\tau) = G_n(\tau, 1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu) + \tau'\omega_n(1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu)$, then $\lambda_n(\tau) = \frac{1}{2}\tau' J\tau + o_p(1)$. Since $\lambda_n(\tau)$ is convex in τ, applying Pollard's convexity lemma (Pollard (1991)):

$$\sup_{\tau\in T}\left|\lambda_n(\tau) - \frac{1}{2}\tau' J\tau\right| \xrightarrow{p} 0,$$

where T is any compact subset of $\mathbb{R}^{r+1}$. Then,

$$G_n(\tau, 1\{\widehat{\kappa}_\nu \geq 0\}\cdot\widehat{\kappa}_\nu) = \frac{1}{2}(\tau - \eta_n)' J(\tau - \eta_n) - \frac{1}{2}\eta_n' J\eta_n + r_n(\tau)$$

with $\sup_{\tau\in T}|r_n(\tau)| = o_p(1)$. So, by Lemma 3 in Buchinsky and Hahn (1998), we have that $\tau_n = \eta_n + o_p(1)$. Therefore, by Lemma A.1

$$n^{1/2}(\widehat{\delta}_\theta - \delta_\theta) \xrightarrow{d} N(0, J^{-1}\Sigma J^{-1}),$$

where $\Sigma = E[\psi\psi']$.

PROOF OF LEMMA A.1 : To prove this lemma we use the assumption that κ_ν is bounded away from zero. This assumption is probably stronger than necessary but it allows us to ignore the trimming using $1\{\widehat{\kappa}_\nu \geq 0\}$, making the asymptotics easier. Assumption (vi) implies that, $K\cdot((K/n_j)^{1/2} + K^{-s}) \to 0$ almost surely for all $j \in \{1, ..., J\}$. Therefore $\sup_{U\in\mathcal{U}}|\widehat{\nu} - \nu_0| = o_p(1)$ (see, e.g., Newey (1997), Theorem 4). Since π_0 is bounded away from zero and one (by (iii)), then $\sup_{U\in\mathcal{U}}|\widehat{\kappa}_\nu - \kappa_\nu| = o_p(1)$. Since κ_ν is bounded away from zero, with probability approaching one the trimming is not binding and we can ignore it for the asymptotics.

$$\omega_n(\widehat{\kappa}_\nu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(1 - \frac{D_i\cdot(1-\widehat{\nu}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i)\cdot\widehat{\nu}_i}{\pi_0(X_i)}\right) + R_n$$

Let π_0^l be the population mean of Z for the l-cell of X and $\widehat{\pi}^l$ its sample counterpart.

$$\begin{aligned} R_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(\frac{(1-D_i)\cdot\widehat{\nu}_i}{\pi_{0i}\cdot\widehat{\pi}_i} - \frac{D_i\cdot(1-\widehat{\nu}_i)}{(1-\widehat{\pi}_i)\cdot(1-\pi_{0i})}\right)\cdot(\widehat{\pi}_i - \pi_{0i}) \\ &= \sum_{l=1}^{L}\left(\frac{1}{\sqrt{n}}\sum_{i_l=1}^{n_l} Z_{i_l} - \pi_0^l\right)\cdot\left(\frac{1}{n_l}\sum_{i_l=1}^{n_l} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})\cdot\widehat{\nu}_{i_l}}{\pi_0^l\cdot\widehat{\pi}^l} - \frac{D_{i_l}\cdot(1-\widehat{\nu}_{i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right)\right). \end{aligned}$$

Also, note that

$$\begin{aligned}
&\left\| \frac{1}{n_l}\sum_{i_l=1}^{n_l} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})\cdot\widehat{\nu}_{i_l}}{\pi_0^l\cdot\widehat{\pi}^l} - \frac{D_{i_l}\cdot(1-\widehat{\nu}_{i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right)\right. \\
&\qquad\qquad \left. - \frac{1}{n_l}\sum_{i_l=1}^{n_l} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})\cdot\nu_{0i_l}}{\pi_0^l\cdot\widehat{\pi}^l} - \frac{D_{i_l}\cdot(1-\nu_{0i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right)\right\| \\
&\qquad = \left\| \frac{1}{n_l}\sum_{i_l=1}^{n} m(U_i)\cdot\left(\frac{(1-D_{i_l})\cdot(\widehat{\nu}_{i_l}-\nu_{0i_l})}{\pi_0^l\cdot\widehat{\pi}^l} + \frac{D_{i_l}\cdot(\widehat{\nu}_{i_l}-\nu_{0i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right)\right\| \\
&\qquad\qquad \leq \sup_{U\in\mathcal{U}}|\widehat{\nu}-\nu_0|\cdot\left\| \frac{1}{n_l}\sum_{i_l=1}^{n} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})}{\pi_0^l\cdot\widehat{\pi}^l} + \frac{D_{i_l}}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right)\right\| = o_p(1).
\end{aligned}$$

Then, applying Lemma 4.3 in Newey and McFadden (1994),

$$\begin{aligned}
&\frac{1}{n_l}\sum_{i_l=1}^{n_l} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})\cdot\widehat{\nu}_{i_l}}{\pi_0^l\cdot\widehat{\pi}^l} - \frac{D_{i_l}\cdot(1-\widehat{\nu}_{i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right) \\
&\qquad = \frac{1}{n_l}\sum_{i_l=1}^{n_l} m(U_{i_l})\cdot\left(\frac{(1-D_{i_l})\cdot\nu_{0i_l}}{\pi_0^l\cdot\widehat{\pi}^l} - \frac{D_{i_l}\cdot(1-\nu_{0i_l})}{(1-\widehat{\pi}^l)\cdot(1-\pi_0^l)}\right) + o_p(1) \\
&\xrightarrow{p} E\left[m(U)\cdot\left(\frac{(1-D)\cdot\nu_0}{(\pi_0(X))^2} - \frac{D\cdot(1-\nu_0)}{(1-\pi_0(X))^2}\right)\middle| X\right] = E\left[m(U)\cdot\left(\frac{(1-D)\cdot Z}{(\pi_0(X))^2} - \frac{D\cdot(1-Z)}{(1-\pi_0(X))^2}\right)\middle| X\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\omega_n(\widehat{\kappa}_\nu) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(1-\frac{D_i\cdot(1-\widehat{\nu}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i)\cdot\widehat{\nu}_i}{\pi_0(X_i)}\right) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} H(X_i)\cdot\{Z_i-\pi_0(X_i)\} + o_p(1).
\end{aligned}$$

To prove,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(1-\frac{D_i\cdot(1-\widehat{\nu}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i)\cdot\widehat{\nu}_i}{\pi_0(X_i)}\right) \\
&\qquad\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(1-\frac{D_i\cdot(1-Z_i)}{1-\pi_0(X_i)} - \frac{(1-D_i)\cdot Z_i}{\pi_0(X_i)}\right) + o_p(1),
\end{aligned}$$

notice that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(U_i)\cdot\left(1-\frac{D_i\cdot(1-\widehat{\nu}_i)}{1-\pi_0(X_i)} - \frac{(1-D_i)\cdot\widehat{\nu}_i}{\pi_0(X_i)}\right) \\
&\qquad\qquad = \sum_{j=1}^{J}\frac{1}{\sqrt{n}}\sum_{i_j=1}^{n_j} m(U_{i_j})\cdot\left(1-\frac{D_{i_j}\cdot(1-\widehat{\nu}_{i_j})}{1-\pi_0(X_{i_j})} - \frac{(1-D_{i_j})\cdot\widehat{\nu}_{i_j}}{\pi_0(X_{i_j})}\right).
\end{aligned}$$

So, we just have to show that for each $j \in \{1, ..., J\}$

$$\frac{1}{\sqrt{n_j}} \sum_{i_j=1}^{n_j} m(U_{i_j}) \cdot \left(1 - \frac{D_{i_j} \cdot (1 - \widehat{\nu}_{i_j})}{1 - \pi_0(X_{i_j})} - \frac{(1 - D_{i_j}) \cdot \widehat{\nu}_{i_j}}{\pi_0(X_{i_j})}\right)$$
$$= \frac{1}{\sqrt{n_j}} \sum_{i_j=1}^{n_j} m(U_{i_j}) \cdot \left(1 - \frac{D_{i_j} \cdot (1 - Z_{i_j})}{1 - \pi_0(X_{i_j})} - \frac{(1 - D_{i_j}) \cdot Z_{i_j}}{\pi_0(X_{i_j})}\right) + o_p(1).$$

This will be done by checking assumptions 6.1 to 6.6 in Newey (1994). Assumptions 6.1 and 6.2 follow directly from the conditions of the theorem (see Newey (1994), page 1373). Assumption 6.3 holds with $d = 0$ and $\alpha_d = s$. Assumption 6.4 holds for $b(z) = 0$ and derivative equal to

$$m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)}\right) (\nu - \nu_0).$$

Assumptions 6.5 and 6.6 follow from: (i) $n_j \cdot K^{-2s} \to 0$; (ii) $K^5/n_j \to 0$ (almost surely). In particular, to check Assumption 6.5 note that (vi) implies that $s > 5/2$, therefore $K \cdot K^{-s} \to 0$ (note that Assumption 6.5 is also valid with $d = 0$). To check assumption 6.6 note that since

$$E\left[\left\| m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)}\right)\right\|^2\right] < \infty,$$

then, there exists a sequence ξ_K such that

$$E\left[\left\| m(U) \cdot \left(\frac{D}{1 - \pi_0(X)} - \frac{1 - D}{\pi_0(X)}\right) - \xi_K p^K(U)\right\|^2\right] \to 0$$

as $K \to \infty$ (see Newey (1994), page 1380 last paragraph.) Now, applying the results in Newey (1994),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \widehat{\nu}_i)}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \widehat{\nu}_i}{\pi_0(X_i)}\right)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - \nu_{0i})}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot \nu_{0i}}{\pi_0(X_i)}\right)$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(U_i) \cdot \left(\frac{D_i}{1 - \pi_0(X_i)} - \frac{1 - D_i}{\pi_0(X_i)}\right) \cdot (Z_i - \nu_{0i}) + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(U_i) \cdot \left(1 - \frac{D_i \cdot (1 - Z_i)}{1 - \pi_0(X_i)} - \frac{(1 - D_i) \cdot Z_i}{\pi_0(X_i)}\right) + o_p(1)$$

and the result of the lemma holds.

Proof of Lemma A.2 : Note that $\rho_n(U_i, 1\{\widehat{\kappa}_\nu \geq 0\} \cdot \widehat{\kappa}_\nu, \tau) - \rho_n(U_i, \kappa_\nu, \tau) = (1\{\widehat{\kappa}_\nu \geq 0\} \cdot \widehat{\kappa}_\nu - \kappa_\nu) \cdot S_n(U_i, \tau)$, where

$$\begin{aligned} S_n(U_i, \tau) &= \theta \cdot [(\epsilon_{\theta i} - n^{-1/2} W_i'\tau)^+ - \epsilon_{\theta i}^+] \\ &+ (1 - \theta) \cdot [(\epsilon_{\theta i} - n^{-1/2} W_i'\tau)^- - \epsilon_{\theta i}^-] + \tau'\zeta_n(U_i), \end{aligned}$$

so $|S_n(U_i, \tau)| \leq n^{-1/2} 1\{|\epsilon_{\theta i}| < n^{-1/2}|W_i'\tau|\} \cdot |W_i'\tau|$. Also,

$$E[n \cdot |S_n(U_i, \tau|] \leq n^{1/2} E[1\{|\epsilon_\theta| < n^{-1/2}|W'\tau|\} \cdot |W'\tau|]$$
$$= E\left[\frac{F_{\epsilon_\theta|W}(n^{-1/2}|W'\tau|) - F_{\epsilon_\theta|W}(-n^{-1/2}|W'\tau|)}{n^{-1/2}} \cdot |W'\tau|\right] \to 2 \cdot E[f_{\epsilon_\theta|W}(0) \cdot |W'\tau|^2] < \infty.$$

Then,

$$
\begin{aligned}
\left|\sum_{i=1}^{n} \rho_n(U_i, 1\{\widehat{\kappa}_\nu \geq 0\} \cdot \widehat{\kappa}_\nu, \tau) - \rho_n(U_i, \kappa_\nu, \tau)\right| &\leq \sum_{i=1}^{n} |1\{\widehat{\kappa}_\nu \geq 0\} \cdot \widehat{\kappa}_\nu - \kappa_\nu| \cdot |S_n(U_i, \tau)| \\
&\leq \sup_{U \in \mathcal{U}} |\widehat{\kappa}_\nu - \kappa_\nu| \cdot \frac{1}{n} \sum_{i=1}^{n} n \cdot |S_n(U_i, \tau)| = o_p(1).
\end{aligned}
$$

Also, by cancellation of cross-product terms,

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} \rho_n(U_i, \kappa_\nu, \tau) - E[\rho_n(U, \kappa_\nu, \tau)]\right)^2\right] &= \sum_{i=1}^{n} E[(\rho_n(U, \kappa_\nu, \tau))^2] \\
&\leq E[1\{|\epsilon_{\theta i}| < n^{-1/2}|W_i'\tau|\} \cdot |W_i'\tau|^2] \\
&\to 0,
\end{aligned}
$$

and the result of the lemma holds.

Proof of Theorem 3.2:

Consistency of $\widehat{\Sigma}$ is easy to prove and we will focus on $\widehat{J}$. By (ii) and (iii), for $0 \leq \epsilon^* \leq \epsilon_\theta$:

$$
\begin{aligned}
E\left[\varphi_h(\delta_\theta) | W, D_1 > D_0\right] &= \int \frac{1}{h} \varphi\left(\frac{y - W'\delta_\theta}{h}\right) f_{Y|W, D_1 > D_0}(y)\, dy \\
&= \int \varphi(z)\, f_{\epsilon_\theta|W, D_1 > D_0}(h \cdot z)\, dz \\
&= f_{\epsilon_\theta|W, D_1 > D_0}(0) + h \cdot \int z \cdot \varphi(z) \left(\frac{\partial f_{\epsilon_\theta|W, D_1 > D_0}(\epsilon^*)}{\partial z}\right) dz \\
&= f_{\epsilon_\theta|W, D_1 > D_0}(0) + O(h). \qquad \text{(A.1)}
\end{aligned}
$$

Therefore,

$$
\lim_{h \to 0} E\left[\varphi_h(\delta_\theta) | W, D_1 > D_0\right] = f_{\epsilon_\theta|W, D_1 > D_0}(0).
$$

By equation (A.1) and condition (iv) in Theorem 3.1, $E[\varphi_h(\delta_\theta)|W, D_1 > D_0]$ is eventually bounded (in absolute value) by a constant. Since W is also bounded, we have that

$$
\begin{aligned}
\lim_{h \to 0} E\left[\kappa_\nu \cdot \varphi_h(\delta_\theta) \cdot WW'\right] &= \lim_{h \to 0} E\left[E\left[\varphi_h(\delta_\theta)|W, D_1 > D_0\right] \cdot WW' | D_1 > D_0\right] \cdot P(D_1 > D_0) \\
&= E\left[f_{\epsilon_\theta|W, D_1 > D_0}(0) \cdot WW' | D_1 > D_0\right] \cdot P(D_1 > D_0).
\end{aligned}
$$

Also, since κ_ν, $\varphi(\cdot)$ and W are bounded,

$$
\operatorname{var}(\kappa_\nu \cdot \varphi_h(\delta_\theta) \cdot WW') = O(1/h^2).
$$

Since $n \cdot h^2 \to \infty$, then

$$
\frac{1}{n} \sum_{i=1}^{n} \kappa_\nu(U_i) \cdot \varphi_{h,i}(\delta_\theta) \cdot W_i W_i' \xrightarrow{p} E\left[f_{\epsilon_\theta|W, D_1 > D_0}(0) \cdot WW' | D_1 > D_0\right] \cdot P(D_1 > D_0). \qquad \text{(A.2)}
$$

Notice that, since κ_ν is bounded away from zero (uniformly in U),

$$
\begin{aligned}
&\left\|\frac{1}{n} \sum_{i=1}^{n} \widehat{\kappa}_\nu(U_i) \cdot \varphi_{h,i}(\widehat{\delta}_\theta) \cdot W_i W_i' - \kappa_\nu(U_i) \cdot \varphi_{h,i}(\widehat{\delta}_\theta) \cdot W_i W_i'\right\| \\
&\leq C \cdot \sup_{U \in \mathcal{U}} |\widehat{\kappa}_\nu - \kappa_\nu| \cdot \frac{1}{n} \sum_{i=1}^{n} \left\|\kappa_\nu \cdot \varphi_{h,i}(\widehat{\delta}_\theta) \cdot W_i W_i'\right\| = C \cdot \sup_{U \in \mathcal{U}} |\widehat{\kappa}_\nu - \kappa_\nu| \cdot \frac{1}{n} \sum_{i=1}^{n} \kappa_\nu \cdot \varphi_{h,i}(\widehat{\delta}_\theta)
\end{aligned}
$$

Under the assumptions of Theorem 3.1, $\sup_{U \in \mathcal{U}} |\widehat{\kappa}_\nu - \kappa_\nu| \xrightarrow{p} 0$. In addition,

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} \kappa_\nu(U_i)\cdot\varphi_{h,i}(\widehat{\delta}_\theta) - \kappa_\nu(U_i)\cdot\varphi_{h,i}(\delta_\theta)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\kappa_\nu(U_i)\cdot\left|\varphi_{h,i}(\widehat{\delta}_\theta) - \varphi_{h,i}(\delta_\theta)\right| \\ &\leq C\cdot\frac{1}{n\cdot h}\sum_{i=1}^{n}\frac{1}{h}\|\widehat{\delta}_\theta - \delta_\theta\|\cdot\|W_i\| \\ &\leq C\cdot n^{1/2}\,\|\widehat{\delta}_\theta - \delta_\theta\|\cdot(n^{1/2}h^2)^{-1} = o_p(1).\end{aligned}$$

As shown above,

$$\frac{1}{n}\sum_{i=1}^{n}\kappa_\nu(U_i)\cdot\varphi_{h,i}(\delta_\theta) = O_p(1).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\kappa}_\nu(U_i)\cdot\varphi_{h,i}(\widehat{\delta}_\theta)\cdot W_iW_i' = \frac{1}{n}\sum_{i=1}^{n}\kappa_\nu(U_i)\cdot\varphi_{h,i}(\widehat{\delta}_\theta)\cdot W_iW_i' + o_p(1). \tag{A.3}$$

By (i) and (iv), for some constant C

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\kappa_\nu(U_i)\cdot\varphi_{h,i}(\widehat{\delta}_\theta)\cdot W_iW_i' - \kappa_\nu(U_i)\cdot\varphi_{h,i}(\delta_\theta)\cdot W_iW_i'\right\| \leq C\cdot n^{1/2}\,\|\widehat{\delta}_\theta - \delta_\theta\|\cdot(n^{1/2}h^2)^{-1} = o_p(1). \tag{A.4}$$

Combining equations (A.2), (A.3) and (A.4), we get $\widehat{J} \xrightarrow{p} J$.

References

ABADIE, A. (1999a): "Bootstrap Tests for the Effect of a Treatment on the Distribution of an Outcome Variable," Mimeo, MIT.

ABADIE, A. (1999b): "Semiparametric Estimation of Instrumental Variable Models for Causal Effects," Mimeo, MIT.

AMEMIYA, T. (1982): "Two Stage Least Absolute Deviations Estimators," *Econometrica* 50, 689-711.

ANGRIST J. D. AND G. W. IMBENS (1991): "Sources of Identifying Information in Evaluation Models," NBER Technical Working Paper No. 117.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444-472.

ATKINSON, A. B. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244-263.

BARRODALE, I., AND F. D. K. ROBERTS (1973): "An Improved Algorithm for Discrete l_1 linear approximation," *SIAM Journal on Numerical Analysis*, 10, 839-848.

BASSETT, G., AND R. KOENKER (1982): "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, 77, 407-415.

BLOOM, H. S., L. L. ORR, S. H. BELL, G. CAVE, F. DOOLITTLE, W. LIN AND J. M. BOS (1997): "The Benefits and Costs of JTPA Title II-A Programs," *The Journal of Human Resources*, 32, 549-576.

BUCHINSKY, M. (1991): *The Theory and Practice of Quantile Regression*, Ph.D. Dissertation. Harvard University.

BUCHINSKY, M. (1994): "Changes in the US Wage Structure 1963-87: Application of Quantile Regression," *Econometrica*, 62, 405-458.

BUCHINSKY, M. (1995): "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study," *Journal of Econometrics* 68, 303-338.

BUCHINSKY, M., AND J. HAHN (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653-671.

CARD, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica* 64, 957-980.

CHAMBERLAIN, G. (1991): "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics Sixth World Congress* 1, ed. by C.A. Sims. Cambridge: Cambridge University Press.

CHARNES A., AND W. W. COOPER (1957): "Nonlinear Power of Adjacent Extreme Point Methods in Linear Programming," *Econometrica* 25, 132-153.

DINARDO, J., NICOLE M. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica* 64, 1001-1045.

FITZENBERGER, B. (1997a): "A Guide to Censored Quantile Regression," in *Handbook of Statistics* 15, ed. by G.S. Maddala and C.R. Rao. New York: Elsevier.

FITZENBERGER, B. (1997b): "Computational Aspects of Censored Quantile Regression," in *Proceedings of the 3rd International Conference on Statistical Data Analysis Based on the L_1-Norm and Related Methods* 31, ed. by Y. Dodge. Haywood California: Institute of Mathematical Statistics Lecture Notes-Monograph Series.

FREEMAN, R. (1980): "Unionism and the Dispersion of Wages," *Industrial and Labor Relations Review*, 34, 3-23.

HAUSMAN, J. A., AND W. K. NEWEY (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss,"*Econometrica* 63, 1445-1476.

HECKMAN, J. AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman and B. Singer. New York: Cambridge University Press.

HECKMAN, J., J. SMITH, AND C. TABER (1994): "Accounting for Dropouts in Evaluations of Social Experiments," NBER technical Working Paper No. 166.

HECKMAN J., J. SMITH AND N. CLEMENTS (1997): "Making the Most Out of Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487-535.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

IMBENS, G. W., AND D. B. RUBIN (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *Review of Economic Studies*, 64, 555-574.

KATZ, L., AND K. MURPHY (1992): "Changes in Relative Wages: Supply and Demand Factors," *Quarterly Journal of Economics*, 107, 35-78.

KING, MERVYN A. (1983): "An Index of Inequality: With Applications to Horizontal Equity and Social Mobility," *Econometrica*, 51, 99-115.

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33-50.

KOENKER, R., AND V. D'OREY (1987): "Computing Regression Quantiles," *Journal of the Royal Statistical Society, Applied Statistics*, 36, 383-393.

LALONDE, R.J. (1995): "The Promise of Public-Sector Sponsored Training Programs," *Journal of Economic Perspectives* 9, 149-168.

LALONDE, R. J., G. MARSCHKE AND K. TROSKE (1996): "Using Longitudinal Data on Establishments to Analize the Effects of Union Organizing Campaigns in the United States," *Annales d'Économie et de Statistique*, 41/42, 155-185.

LEWIS, H. G. (1986): *Union Relative Wage Effects: A Survey.* Chicago: University of Chicago Press.

MANSKI, C. F. (1988): *Analog Estimation Methods in Econometrics.* New York: Chapman and Hall.

MANSKI, C. F. (1994): "The Selection Problem," in *Advances in Econometrics Sixth World Congress* 2, ed. by C.A. Sims. Cambridge: Cambridge University Press.

MANSKI, C. F. (1997): The Mixing Problem in Programme Evaluation, Review of Economic Studies 64, 537-553.

NEWEY, W. K, AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* 4, ed. by R. F. Engle and D. McFadden. Amsterdam: Elsevier Science Publishers.

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica* 62, 1349-1382.

NEWEY, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79, 147-168.

NEWEY, W. K, AND J. L. POWELL (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory* 6, 295-317.

ORR, L. L., H. S. BLOOM, S. H. BELL, F. DOOLITTLE, W. LIN AND G. CAVE (1996): *Does Training for the Disadvantaged Work?*, Washington, DC: The Urban Institute.

POLLARD, D. (1991): "Asymptotics for Least Absolute Deviation Estimators," *Econometric Theory*, 7, 186-199.

PORTNOY, S. L. (1991): "Asymptotic Behavior of Regression Quantiles in Non-stationary Dependent Cases," *Journal of Multivariate Analysis*, 38, 100-113.

POWELL, J. L. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569-1575.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688-701.

RUBIN, D. B. (1977): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1-26.

US DEPARTMENT OF LABOR (1995): What's Working (and What's Not), A Summary of Research on the Economic Impacts of Employment and Training Programs, Washington, DC: US Government Printing Office, January.

US DEPARTMENT OF LABOR (1999): Employment and Training Administration, Job Training Partnership Fact Sheet, http://www.doleta.gov/programs/factsht/jtpa.htm.

Table I
Means and Standard Deviations

	Entire Sample	Assignment: Treatment	Assignment: Control	Difference (t-stat.)
A. Men				
Number of observations	5,102	3,399	1,703	
Baseline Characteristics				
Age	32.91 [9.46]	32.85 [9.46]	33.04 [9.45]	-.19 (-.67)
High school or GED	.69 [.45]	.69 [.45]	.69 [.45]	-.00 (-.12)
Married	.35 [.47]	.36 [.47]	.34 [.46]	.02 (1.64)
Black	.25 [.44]	.25 [.44]	.25 [.44]	.00 (.04)
Hispanic	.10 [.30]	.10 [.30]	.09 [.29]	.01 (.70)
Worked less than 13 weeks in past year	.40 [.47]	.40 [.47]	.40 [.47]	.00 (.56)
Experimental Characteristics				
Second follow-up	.29 [.46]	.30 [.46]	.28 [.45]	.02 (1.14)
Training	.42 [.49]	.62 [.48]	.01 [.11]	.61 (70.34)
Service strategy:				
Classroom training	.20 [.40]	.21 [.41]	.19 [.39]	.02 (1.73)
OJT/JSA	.50 [.50]	.50 [.50]	.50 [.50]	.00 (.07)
Other	.29 [.46]	.29 [.45]	.31 [.46]	-.02 (-1.57)
Outcome variable:				
30 month earnings	19,147 [19,540]	19,520 [19,912]	18,404 [18,760]	1,116 (1.96)

continued from previous page

	Entire Sample	Assignment: Treatment	Assignment: Control	Difference (t-stat.)
B. Women				
Number of observations	6,102	4,088	2,014	
Baseline Characteristics				
Age	33.33	33.33	33.35	-.02
	[9.78]	[9.77]	[9.81]	(-.09)
High school or GED	.72	.73	.70	.03
	[.43]	[.43]	[.44]	(2.01)
Married	.22	.22	.21	.01
	[.40]	[.40]	[.39]	(1.55)
Black	.26	.27	.26	.01
	[.44]	[.44]	[.44]	(.95)
Hispanic	.12	.12	.12	-.00
	[.32]	[.32]	[.33]	(-.89)
Worked less than 13 weeks in past year	.52	.52	.52	-.00
	[.47]	[.47]	[.47]	(-.08)
AFDC	.31	.30	.31	-.01
	[.46]	[.46]	[.46]	(-1.03)
Experimental Characteristics				
Second follow-up	.26	.26	.25	.01
	[.44]	[.44]	[.43]	(.45)
Training	.45	.66	.02	.64
	[.50]	[.47]	[.13]	(80.24)
Service strategy:				
Classroom training	.38	.38	.39	-.01
	[.49]	[.49]	[.49]	(-.37)
OJT/JSA	.37	.37	.38	-.01
	[.48]	[.48]	[.49]	(-.85)
Other	.24	.25	.23	.02
	[.43]	[.43]	[.42]	(1.40)
Outcome variable:				
30 month earnings	13,029	13,439	12,197	1,242
	[13,415]	[13,614]	[12,964]	(3.46)

Note: The first three columns of the table report means and standard deviations (in brackets) for the National JTPA Study 30-month earnings sample. The last column shows the difference in means by assignment status and reports the t-statistic (in parenthesis) for the null hypothesis of equality in means.

Table II
OLS and IV Estimates of Training Impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Note: Columns (1) and (2) show the differences in earnings by training status; columns (3) and (4) show differences by assignment status. Columns (5) and (6) report the result of using assignment status as an instrument for training. The covariates used in columns (2), (5) and (6) are *High school or GED*, *Black*, *Hispanic*, *Married*, *Worked less than 13 weeks in past year*, *AFDC* (for women), plus indicators for the service strategy recommended, age group and second follow-up survey. Robust standard errors in parenthesis.

Table III
Quantile Regression and OLS Estimates

Dependent variable: 30-month earnings

	OLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
A. Men						
Training	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of Training	21.20	135.56	75.20	34.50	17.24	13.43
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	-2,354	-134	-500	-2,084	-3,576	-3,609
	(626)	(194)	(324)	(684)	(1087)	(1,331)
Hispanic	251	91	278	925	-877	-85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked less than 13	-6,582	-1,090	-3,097	-7,610	-9,834	-9,951
weeks in past year	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	-216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)
B. Women						
Training	2,215	367	1,013	2,707	2,729	2,058
	(334)	(105)	(170)	(425)	(578)	(657)
% Impact of Training	18.46	60.76	44.42	32.25	14.47	8.09
High school or GED	3,442	166	681	2,514	5,778	6,373
	(341)	(99)	(156)	(396)	(606)	(762)
Black	-544	22	-60	-129	-866	-1,446
	(397)	(115)	(188)	(451)	(679)	(869)
Hispanic	-1,151	-31	-222	-995	-1,620	-1,503
	(488)	(130)	(194)	(546)	(911)	(992)
Married	-667	-213	-392	-758	-1,048	-902
	(436)	(127)	(209)	(522)	(785)	(970)
Worked less than 13	-5,313	-1,050	-3,240	-6,872	-7,670	-6,470
weeks in past year	(370)	(137)	(289)	(522)	(672)	(787)
AFDC	-3,009	-398	-1,047	-3,389	-4,334	-3,875
	(378)	(107)	(174)	(468)	(737)	(834)
Constant	10,361	649	2,633	8,417	16,498	20,689
	(815)	(255)	(490)	(966)	(1,554)	(1,232)

Note: The table reports OLS and quantile regression estimates of the effect of training on earnings. The specification used also includes indicators for service strategy recommended, age group and second follow-up survey. Robust standard errors in parenthesis.

Table IV
Quantile Treatment Effects and 2SLS Estimates

Dependent variable: 30-month earnings

	2SLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
A. Men						
Training	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of Training	8.55	5.19	11.99	9.64	10.69	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	-2,349	-171	-377	-2,656	-4,182	-3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked less than 13 weeks in past year	-6,575	-1,932	-4,195	-7,009	-9,289	-9,078
	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	-134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)
B. Women						
Training	1,780	324	680	1,742	1,984	1,900
	(532)	(175)	(282)	(645)	(945)	(997)
% Impact of Training	14.60	35.47	23.14	18.37	10.06	7.39
High school or GED	3,470	262	768	2,955	5,518	5,905
	(342)	(178)	(274)	(643)	(930)	(1026)
Black	-554	0	-123	-401	-1,423	-2,119
	(397)	(204)	(318)	(724)	(949)	(1,196)
Hispanic	-1,145	-73	-138	-1,256	-1,762	-1,707
	(488)	(217)	(315)	(854)	(1,188)	(1,172)
Married	-652	-233	-532	-796	38	-109
	(437)	(221)	(352)	(846)	(1,069)	(1,147)
Worked less than 13 weeks in past year	-5,329	-1,320	-3,516	-6,524	-6,608	-5,698
	(370)	(254)	(430)	(781)	(931)	(969)
AFDC	-2,997	-406	-1,240	-3,298	-3,790	-2,888
	(378)	(189)	(301)	(743)	(1,014)	(1,083)
Constant	10,538	984	3,541	9,928	15,345	20,520
	(828)	(547)	(837)	(1,696)	(2,387)	(1,687)

Note: The table reports 2SLS and QTE estimates of the effect of training on earnings. Assignment status is used as an instrument for training. The specification used also includes indicators for service strategy recommended, age group and second follow-up survey. Robust standard errors in parenthesis.